

引用格式: 李雪剑, 秦斌, 肖艺峰, 等. 改进随机森林-蒙特卡罗法在 A 型液舱支座结构可靠性分析中的应用 [J]. 中国舰船研究, 2022, 17(1): 147-153, 165.

LI X J, QIN B, XIAO Y F, et al. An improved random forest-Monte Carlo method and application for structural reliability analysis of A-type independent liquid tank support structure[J]. Chinese Journal of Ship Research, 2022, 17(1): 147-153, 165.

改进随机森林-蒙特卡罗法在 A 型液舱支座结构可靠性分析中的应用



扫码阅读全文

李雪剑*, 秦斌, 肖艺峰, 付泽坤

江南造船(集团)有限责任公司 江南研究院, 上海 201913

摘要: [目的] 随着液化天然气(LNG)船舶结构研究和设计深度的提高, 需要有能够快速和准确地评估不确定性因素的可靠性分析方法。为此, 提出基于改进随机森林-蒙特卡罗(RF-MC)法来解决 A 型独立液舱支座结构失效概率的计算问题。[方法] 首先, 根据不确定性因素的概率分布, 使用 MC 法生成样本集; 然后, 以局部离群因子为准则, 筛选出失效面附近的样本点, 再对筛选出的样本点进行有限元计算后添加至训练集, 通过重复训练随机森林近似模型, 直至满足精度要求; 最后, 使用近似模型判别样本点是否失效, 结合 MC 法计算结构的失效概率。[结果] 综合考虑算法的准确率、复杂度和效率并结合算例 1 和 2, 可以发现在分析可靠性问题时改进 RF-MC 法比 MC 和 BP-MC 等方法具有更大优势。算例 3 的应用结果表明了改进 RF-MC 法在 A 型独立液舱支座结构可靠性分析中的适用性。[结论] 研究结果可为 LNG 船舶的优化设计提供可行的技术方案。

关键词: 结构可靠性; 局部离群因子; 随机森林

中图分类号: U661.42

文献标志码: A

DOI: 10.19693/j.issn.1673-3185.02181

An improved random forest-Monte Carlo method and application for structural reliability analysis of A-type independent liquid tank support structure

LI Xuejian*, QIN Bin, XIAO Yifeng, FU Zekun

Jiangnan Institute of Technology, Jiangnan Shipyard (Group) Co., Ltd., Shanghai 201913, China

Abstract: [Objectives] In response to the increasing depth of research and design on liquefied natural gas (LNG) ship structures, higher requirements are put forward for a reliability analysis method that can quickly and accurately evaluate uncertain factors. This paper proposes a method based on an improved random forest-Monte Carlo method (RF-MC) to solve the calculation of the failure probability of A-type independent liquid tank support structures. [Methods] First, the MC method is used to generate a sample set according to the probability distribution of uncertain factors, then take the local outlier factor (LOF) as the criterion for filtering out sample points near the failure surface. After selecting the sample points, they are calculated using finite element software and added to the training set to train the random forest (RF) model. The generation, filtering and training process is repeated until the approximate model meets the accuracy requirements. Finally, the approximate model is used to determine whether the sample points are invalid, then combined with the MC method to calculate the failure probability of the structure. [Results] Considering the accuracy, complexity and efficiency of the algorithm, and combined with Cases 1 and 2, it is found that the improved RF-MC method has better advantages than MC or biased probability (BP)-MC in analyzing reliability problems. The results of Case 3 show applicability of the method in reliability analysis of an A-type independent liquid tank support structure. [Conclusions] This study provides a feasible technical solution for future optimization design of liquefied gas carriers.

Key words: structural reliability; local outlier factor; random forest

收稿日期: 2020-11-13

修回日期: 2021-01-15

网络首发时间: 2022-01-28 12:02

作者简介: 李雪剑, 男, 1992 年生, 硕士, 工程师

秦斌, 男, 1989 年生, 工程师

肖艺峰, 男, 1995 年生, 助理工程师

付泽坤, 男, 1993 年生, 助理工程师

*通信作者: 李雪剑

0 引言

近年来,随着液化天然气市场的发展,涌现了不同类型的液化天然气(LNG)船,例如,采用C型独立液舱维护系统的 $3 \times 10^4 \text{ m}^3$ (30K)LNG船,A型独立液舱的 8.4×10^4 和 $8.6 \times 10^4 \text{ m}^3$ (84K和86K)大型全冷式液化气船(VLGC),以及B型独立液舱的 $9.8 \times 10^4 \text{ m}^3$ (98K)超大型液化乙烷运输船(VLEC)等。随着液化天然气船品类越来越丰富,船体结构设计也趋于复杂,除了确定性强度要求外,还需对影响结构安全性、适用性、耐久性的不确定因素进行准确评估^[1]。

结构可靠性设计是一种基于数理统计和概率论,用于处理工程中可能存在的各种随机变量的方法。结构设计中不确定的随机变量包括:外力载荷、几何尺寸、材料属性、生产过程和工作环境等,以向量的形式表示为 $\mathbf{X} = (X_1, X_2, \dots, X_n)$ 。在结构可靠性设计中,通过定义一个关于 $\mathbf{X}$ 的功能函数 $Z(\mathbf{X})$ 来明确结构的极限状态, $Z < 0$ 表明结构处于失效状态。实际工程计算中通常采用近似计算的方法获得结构的失效概率,典型的如一阶可靠性法(first-order reliability method, FORM)、二阶可靠性法(second-order reliability method, SORM)和蒙特卡罗法(Monte Carlo, MC)等^[2]。但是,FORM和SORM在处理隐式、失效面附近的非线性程度很高的功能函数时,这两种方法计算的精度难以满足工程要求;而MC法虽然可以求解隐式问题,但在处理复杂系统时,单个样本点的计算时间过长,大量的采样点使得计算成本巨大,在实际工程中不具备可操作性。

为了解决MC法计算效率低的问题,可以采用近似模型技术代替传统的复杂分析模型中的有限元程序,以此来缩短单个样本点的计算时间。金伟良等^[3]通过构建最小二乘支持向量机(LS-SVM)模型拟合失效面,并结合MC法生成的样本点来计算结构的失效概率。李放等^[4]用人工少数类过采样(synthetic minority over-sampling technique, SMOTE)算法合成失效面附近的新样本点,添加到训练集来构建梯度以提升决策树(gradient boosting decision tree, GBDT)近似模型,并完成了结构的可靠性分析。然而,因失效点和可靠点在设计空间中的占比不均衡,直接用MC法生成的样本点作为训练集来构建近似模型,容易造成模型在失效面附近的精度不够。而SMOTE算法构造的新样本是原数据集中不存在的,并不能提供有益信息且会增加类之间出现重叠的可能性。

基于上述问题,本文拟提出以局部离群因子

(local outlier factor, LOF)作为筛选准则,直接从样本集中获得样本分布均衡的训练集,并选择更加容易实现且能够并行计算的随机森林(random forest, RF)模型作为近似模型来判别失效点,用以提高MC法计算结构可靠性的效率。

1 算法原理

1.1 LOF算法

LOF算法是一种典型的基于密度的高精度离群点检测方法^[5],其核心思想是针对给定的数据集,为每个数据点分配一个局部离群因子,用以表征其局部邻域内点的密集程度。该因子越大说明数据点偏离局部中心越远,该点是离群点的可能性也越高,在“类别不平衡”问题中,它就是少数类数据点。以下为LOF算法最重要的几个概念定义。

定义1: k -邻近距离。对象数据点 p 的 k -邻近距离记作 $k\text{-distance}(p)$,其中 k 为任意正整数。对象 o 属于数据集 D ,将对象 o 与 p 之间的距离记作 $d(p, o)$ 。满足以下条件时, $k\text{-distance}(p) = d(p, o)$ 。

- 1) 在集合中至少存在 k 个对象, $o' \in D \setminus \{p\}$,使得 $d(p, o') \leq d(p, o)$;
- 2) 在集合中至多存在 $k-1$ 个对象, $o' \in D \setminus \{p\}$,使得 $d(p, o') < d(p, o)$;

定义2: k -邻域。给定对象 p 的 k -邻近距离,则 p 的 k -邻域是指数据集中与 p 的距离不超过 k -邻近距离的所有对象,记作

$$N_{k\text{-distance}(p)}(p) = \{q \in D \setminus \{p\} | d(p, q) \leq k\text{-distance}(p)\} \quad (1)$$

式中, q 为对象 p 的 k 最近邻。

定义3: 可达距离。对象 o 相对于对象 p 的可达距离定义如下:

$$\text{reach-dist}_k(o, p) = \max\{k\text{-distance}(p), d(p, o)\} \quad (2)$$

定义4: 局部可达密度。对象 p 的局部可达密度是 p 的 k 邻域内所有点平均可达距离的倒数,定义为

$$\text{lrd}_k(p) = 1 / \left(\frac{\sum_{o \in N_{k\text{-distance}(p)}(p)} \text{reach-dist}_k(o, p)}{|N_{k\text{-distance}(p)}(p)|} \right) \quad (3)$$

对于同一个簇内邻域点,它们的可达距离可能是较小值,使得 $N_{k\text{-distance}(p)}(p)$ 内点的可达距离之和越小,局部可达密度越大。

定义5: 局部离群因子。对象 p 的局部离群因

子是邻域 $N_{k-distance(p)}(p)$ 内其他点与 p 的局部可达密度之比的平均数,即

$$LOF_k(p) = \frac{\sum_{o \in N_{k-distance(p)}(p)} \frac{lrd_k(o)}{lrd_k(p)}}{|N_{k-distance(p)}(p)|}$$

(4)

对于 $LOF_k(p)$,若该值接近 1,说明 p 的邻域点密度差不多, p 可能和邻域属于同一个簇。若小于 1,说明 p 的密度高于其邻域点密度,表明 p 处在一个相对密集的区域。而若大于 1,则表明 p 与其他数据点比较疏远, p 很可能是离群点。

LOF算法通过求出数据集中每个点的局部离群因子,并对这些 $LOF_k(p)$ 值进行排序,然后设定阈值 δ ,将排序后的离群因子值超过 δ 的对象作为离群点集合,其中 δ 值根据不同数据集规模和设计经验确定。

1.2 随机森林模型

随机森林^[6]是并行式集成学习方法套袋法(Bagging)^[7]的一个扩展变体,它不仅容易实现而且计算量小。RF 在基学习器的训练过程中引入随机属性的选择,然后通过 Bagging 对基学习器进行集成。通常采用分类回归树,即 CART 决策树作为基学习器,因为它能够有效地处理数据样本的非线性问题,适用于解决数据样本和目标变量之间关系模糊不明确的复杂性问题^[8]。

CART 决策树的生成就是递归地构建二叉树,并且针对分类和回归问题使用不同的策略。其中分类树使用基尼指数(Gini 不纯度)最小化准则选择属性构建模型,基尼指数是数据集不确定

性的度量,其定义如下:

$$Gini(p) = \sum_{i=1}^S p_i(1 - p_i) = 1 - \sum_{i=1}^S p_i^2$$

(5)

式中: S 为数据集包含的类别数; p_i 为数据点属于第 i 类的概率。

回归树采用启发式方法对属性空间进行划分,划分的每个单元都对应 1 个固定值,可使用平方误差最小化准则求解固定值的最优解,它的定义为

$$f(x) = \sum_{m=1}^M c_m I(x \in R_m)$$

(6)

式中: R_m 为划分的第 m 个单元,其中 $m = 1, 2, \dots, M$, M 为划分的最大单元数; c_m 为单元 R_m 上的固定输出值; $I(x \in R_m)$ 为指示函数,当 x 属于单元 R_m 时值为 1,否则为 0。

RF 模型构建的步骤是:首先,从给定的含有 n 个训练样本、 m 个属性的数据集中进行 Bootstrap 有放回的抽样,组成含 n 个样本的新训练集 S_i ;然后,以训练集 S_i 训练模型节点时,从节点所有的样本属性中随机选择 m' 个属性,选择其中最佳属性分裂节点来构建 CART 树;重复上述步骤 T 次得到 T 个决策树;这 T 个 CART 树形成随机森林,通过结合策略将每棵树的预测结果进行整合后作为 RF 模型的输出。RF 模型原理如图 1 所示,图中 T_i 为第 i 个决策树模型。

通过样本扰动和属性扰动使基学习器具有多样性,最终集成的 RF 模型的泛化性能由于个体学习器之间差异度的增加而进一步提升,不仅准

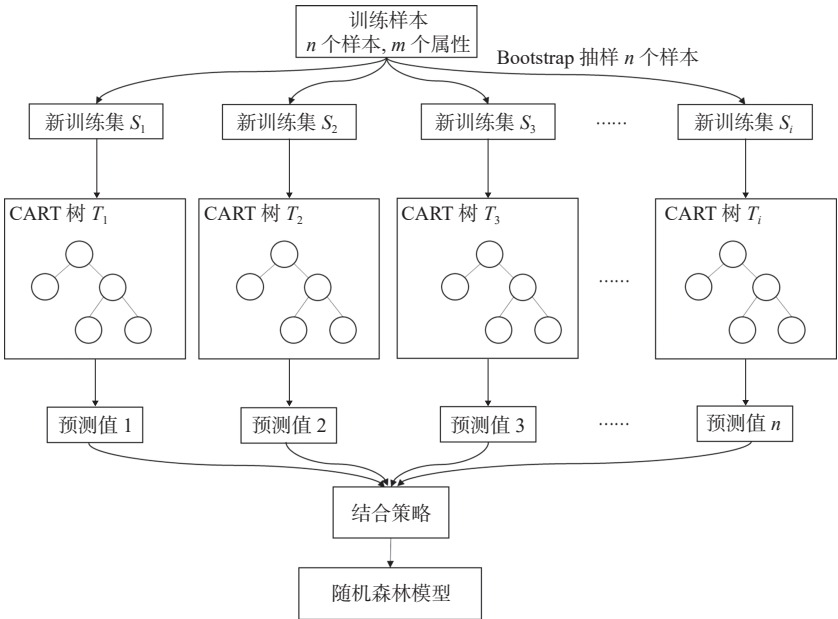


图 1 随机森林模型原理

Fig. 1 Principle of RF model

率、运行速度要高于 Adaboost 算法^[9], 同时对于离群点的鲁棒性也更好。而且, Bootstrap 抽样使得生成每个决策树时约有 1/3 的样本(称为袋外样本)未被使用, 通过计算袋外误差率可以建立模型误差的无偏估计, 不再需要独立的测试集, 使得模型的构建更加简单。

2 基于改进 RF-MC 的结构可靠性分析

实际工程中事物的不确定性现象是客观存在的, 对于不确定性因素可用随机性来描述^[10-11]。因此, 在结构可靠性分析中可将结构载荷、材料强度、弹性模量等视为随机变量^[12]。随机变量通常具有相应的统计规律, MC 法就是根据统计规律随机采样来估算结构的近似失效概率^[13]:

$$\hat{l} = \frac{1}{N} \sum_{i=1}^N I[Z(X_i) < 0] \quad (7)$$

式中: N 为总采样点数; X_i 为样本点; $Z(X_i)$ 为结构的功能函数。式(7)以失效点在所有样本点中出现的频率作为结构的失效概率。实际工程结构的失效概率 l 非常小, 使用 MC 法估算的相对误差为

$$e = \frac{\sqrt{\text{Var}(\hat{l})}}{E[\hat{l}]} = \sqrt{\frac{1-l}{Nl}} \approx \sqrt{\frac{1}{Nl}} \quad (8)$$

式中: $\text{Var}(\hat{l})$ 为近似失效概率的方差; $E[\hat{l}]$ 为近似失效概率的均值。

假设结构的实际失效概率为 $l = 10^{-4}$, 通常估算的相对误差要满足 $e = 0.01$, 为了保证近似精度, 根据式(8), MC 法需要的总采样数应达到 $N \approx \frac{1}{e^2 l} = 10^8$ 。由于 $Z(X_i)$ 常常不具有明确合理的解析表达式, 因此需要借助计算耗时的有限元软件来分析样本点是否失效。而总采样点数 $N = 10^8$, 若通过有限元法判别每个样本点是否失效, 将耗时巨大。

本文提出的改进 RF-MC 法旨在利用计算量小、高效并行计算的 RF 模型替代有限元法来判断失效点。通过 MC 法对高可靠性的结构进行采样时, 其所获得的样本集极不均衡, 失效点占比非常小, 属于少数类样本^[14]。当使用样本不平衡的数据集对 RF 模型进行训练时, 模型的预测精度可能会很低。由于少数类样本也可以被看作离群点或异常点, 因此, 本文使用 LOF 算法进行样本筛选, 以保证训练集中的可靠点和失效点的比例均衡, 从而提高模型在失效面附近的精度。改进 RF-MC 算法的流程如图 2 所示。

1) 确定功能函数 $Z(X_i)$ 中影响结构状态和性

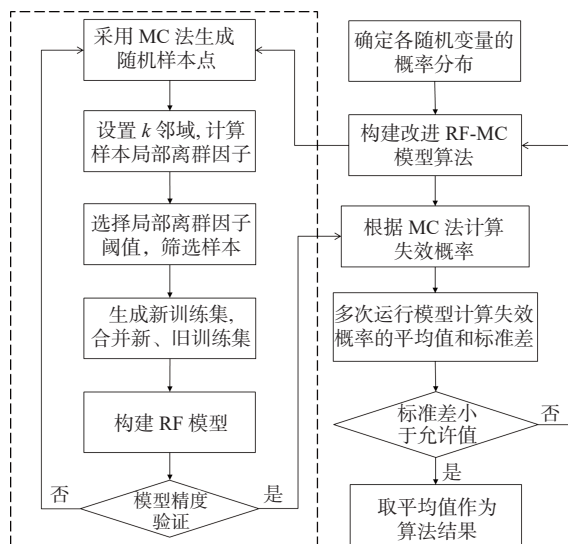


图 2 改进 RF-MC 方法流程

Fig. 2 Workflow of improved RF-MC method

质的随机变量及其所服从的概率分布。对于概率分布不确定的变量, 可以使用正态分布来近似, 均值取设计中的实际值, 变异系数取 0.1。

2) 采用 MC 法在设计空间内生成含有 N 个样本点 $X_i (i = 1, 2, \dots, N)$ 的样本集 X 。此时, 无需调用有限元模型计算样本点的功能函数值。

3) 设置邻域值 k , 使用 LOF 算法计算样本集中每个点的局部离群因子, 即 $LOF_k(p)$ 。因为 $LOF_k(p)$ 反映了样本邻域点的分布密度, 所以当 $LOF_k(p) > 1$ 时, 表明样本点 X_i 离其他点比较疏远。在工程结构中, 失效域内的样本点出现概率非常低, 在样本空间中会远离其他点, 亦即其 $LOF_k(p) \gg 1$, 因此, 可以将失效点看作离群点通过 LOF 算法筛选得到。

4) 确定局部离群因子的阈值上限 δ_u 和阈值下限 δ_l , 将样本集 X 中 $\delta_l \leq LOF_k(p) \leq \delta_u$ 的样本点加入到子集 X' 。通过 δ_l 筛选出样本集中的离群点, 离群点不仅包含失效面附近的点, 而且还包括远离失效面的点, 这些属于噪声的点会增加模型的误差, 因此, 可使用 δ_u 来排除远离失效面的点。

5) 调用有限元模型计算样本子集 X' 的功能函数值作为响应集 $Z(X')$, 然后将 X' 作为输入, $Z(X')$ 作为输出建立 RF 模型。选取合适的模型参数可提高拟合精度, 但参数值与精度值通常不是线性关系, 若 CART 树个数过大或过小都不利于模型精度, 因此使用机器学习库 scikit-learn 提供的交叉验证和网格搜索模块选取合适的 RF 模型参数。

6) 使用袋外样本评估 RF 模型的平均绝对百分比误差 ϵ 。若模型精度不符合要求, 返回步骤 2) 重新训练 RF 模型, 并将 $(X', Z(X'))$ 添加至新生

成的训练集。本文验证所允许的误差 $\epsilon < 0.05$ 。

7) 重复上述迭代过程,直至模型精度满足平均绝对百分比误差的要求。使用 RF 模型替代有限元法预测每个样本点的功能函数值,并根据 MC 法的公式(5)计算失效概率。

8) 考虑到算法结果存在一定随机性,在局部离群因子和 RF 模型参数确定后将多次计算结构失效概率,若所有结果的标准差小于允许值,则将平均值作为算法最终结果,反之,则说明算法不稳定、拟合效果不佳,需要重新调整参数训练模型。本文建议的运行次数是 9 次,标准差的允许值取平均值的 10%。

本文提出的改进 RF-MC 法主要通过局部离群因子的邻域值 k , δ_0 和 δ_1 来控制样本点的筛选。为了使生成的训练集中的可靠点和失效点分布均衡,合理控制参数很重要。考虑到工程中随机变量服从的概率分布大多未知,所以本文方法采用的正态分布,如此,可以使用服从多元正态分布的样本点来研究控制参数对算法效率的影响。

这里,分析含两个设计变量的可靠性问题,其变量相互独立且服从标准的正态分布,且使用 LOF 算法检测样本集中的离群点。改变邻域值 k 并固定其他参数来观察离群点数量变化,如图 3 所示,可以发现离群点数量先减后增。考虑到正态分布的特征,较小的 k 值计算得到的样本点 $LOF_k(p)$ 比较接近,难以区分离群点和密集点,随着 k 值的增大,算法区分出离群点的能力增强,使得曲线呈下降趋势;而当 k 值过大时,失效面附近的可靠点因既不靠近样本点的密集区域,又远离失效点,计算得到的 $LOF_k(p)$ 较大时也会被视为离群点,使曲线呈上升趋势。因此,较大的 k 值有助于获得分布均衡的训练集,但 k 值过大会增加计算量。所以综合考虑后,建议 k 值取 30。本文认为阈值超过 1.5 的点是噪声,故要求 $\delta_0 \leq 1.5$,而 δ_1 需根据数据集的规模和设计经验选取,一般取值范围为 [1,1.3]。当局部离群因子较接近时,可根据 $LOF_k(p)$ 进行排序,取值最大的一部分样本点。为简化流

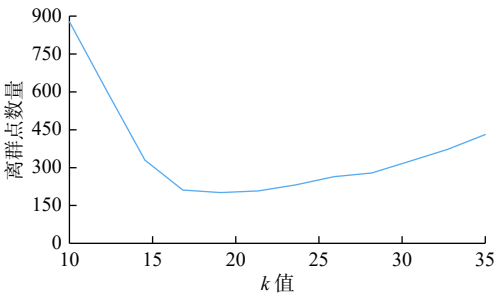


图 3 离群点数量随 k 值变化曲线
Fig. 3 The number of outliers varying with k -value

程,本文取排序后前 50 个点加入训练集。

3 算例验证

3.1 算例 1

为了对比本文方法与陈松坤等^[14]提出的 BP-MC 法,本文采用该文献中的一个具有解析表达式的功能函数:

$$G(U_1, U_2) = 0.5(U_1 - 2)^2 - 1.5(U_2 - 5)^3 - 3 \tag{9}$$

式中: U_1 和 U_2 均为设计变量,两者相互独立且均服从标准正态分布。分别使用 MC 法、BP-MC 法和改进 RF-MC 法进行求解,同时,为了保证结果具有可比性,使用了相同数量的样本训练模型,结果如表 1 所示。

表 1 算例 1 计算结果
Table 1 Calculation results of Case 1

方法	失效概率 P_f	相对误差/%	样本数量
MC	2.85×10^{-5}	—	5×10^7
BP-MC	3.22×10^{-5}	12.98	124
改进 RF-MC	3.15×10^{-5}	4.91	124

综合考虑计算的准确率和样本点计算次数,可以发现:改进 RF-MC 法使用的样本点数量远小于 MC 法,计算时间得到极大缩减;与 BP-MC 法相比,改进 RF-MC 法准确率更高,且在筛除远离失效面的点时直接使用 $LOF_k(p)$,而 BP-MC 法需要训练出模型进行预测后筛选样本,因此改进 RF-MC 法更加简便。

3.2 算例 2

平面十杆桁架结构被广泛用于验证结构设计优化算法。该算例^[1]是失效面为隐式的十杆桁架结构,如图 4 所示,图中, L 为桁架长度, P 为节点上作用的载荷。由图可见,节点(1)和节点(6)为简支,节点(2)和节点(3)均含垂直向下的集中力(444.822 kN)。水平、横向和对角桁架的横截

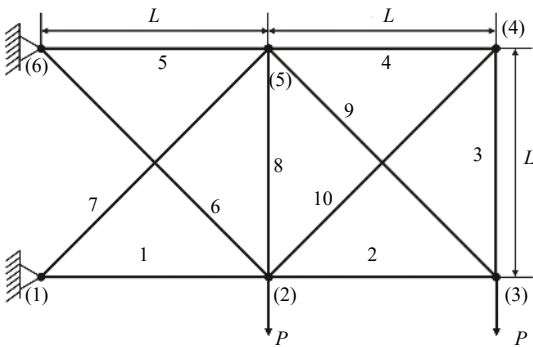


图 4 十杆桁架结构
Fig. 4 Crossbar truss structure

面积 A_1 , A_2 和 A_3 为随机变量均服从正态分布, 参数见表 2。

表 2 基本随机变量统计特征值
Table 2 Statistical characteristic values of basic random variable

随机变量/cm ²	均值 μ	标准差 σ	分布类型
A_1	83.87	8.387	正态
A_2	12.90	1.290	正态
A_3	58.06	5.806	正态

除表 2 中参数外, 其他参数是确定的。包括: $L=9.144$ m; 弹性模量 $E=68.95$ MPa; 密度 $\rho=2.7\times 10^3$ kg/m³; 集中力 $P=444.822$ kN。由于算例具有隐式的失效面, 需要使用有限元分析结构响应来计算功能函数值。假设节点 (3) 的最大许用位移为 101.6 mm, 有限元计算得到的位移为 $d(A_1, A_2, A_3)$, 则节点 (3) 的功能函数为

$$Z = 101.6 - d(A_1, A_2, A_3) \tag{10}$$

通过对比改进 RF-MC 法与工程中被普遍视为理论解的 MC 法, 可以发现改进 RF-MC 法在复杂结构可靠性分析中具有较高的精度, 结果如表 3 所示。

表 3 算例 2 计算结果
Table 3 Calculation results of Case 2

方法	失效概率 P_f	相对误差/%	样本数量
MC	0.176 40	—	10^5
改进RF-MC	0.179 85	1.96	200

3.3 算例3

本文以某型采用 A 型独立液舱的液化石油气 (LPG) 船中间舱某一档垂向支座的支撑肘板为研究对象, 使用改进 RF-MC 法计算其结构可靠性, 并与采用 MC 法得到的结构失效概率比较, 图 5 是采用 MC 法计算的失效概率随采样数的收

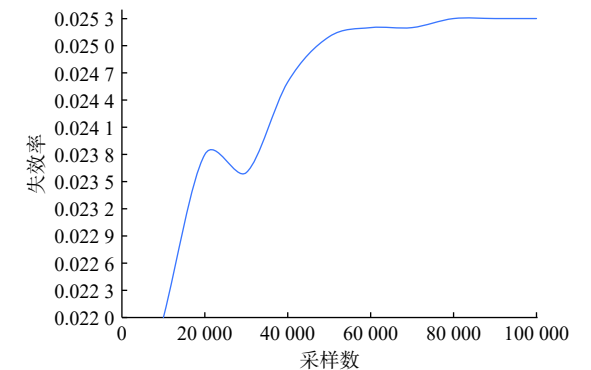


图 5 MC 法计算结果
Fig. 5 Calculation result of MC method

敛情况。为了验证算法的准确性, 需要通过有限元软件计算获取模型训练数据。使用 Abaqus 有限元软件, 借助于该软件内核支持的 Python 语言进行编程修改 input 文件的设计参数来获取样本点。

选取的模型范围是中间舱顶边舱以下左舷结构, 如图 6 所示。有限元模型共有 9 715 个单元, 其中板材使用壳单元, 加强筋使用梁单元建模, 边界条件是模型两端固支, 载荷是目标支座及附近支座的支反力, 它是根据船级社软件计算出的所有工况中的最大作用力。通过分析选取出 21 个对支撑肘板强度有影响的设计变量, 分别为: 11 个支座支反力 (包括垂向支座、防横摇支座、防纵摇支座), 8 个板材厚度 (包括目标支座肘板、内底板、纵桁、肋板等) 和 2 个纵骨腹板厚度 (包括内底纵骨、外底板纵骨)。设计变量服从的分布如表 4 所示。需要说明的是, 当设计变量服从均匀分布时, 表 4 中参数 1 为上限, 参数 2 为下限; 当设计变量服从正态分布时, 参数 1 为均值, 参数 2 为方差。表 4 中各变量含义分别为: f_1 为防纵摇支座支反力; f_2 为防横摇支座支反力; f_3 为垂向支座支反力; d_1 为纵桁板厚; d_2 为垂向支座肘板厚度; d_3 为防纵摇支座肘板厚度; d_4 为肋板厚度; d_5 为支座支撑结构板厚; d_6 为内底板厚度; d_7 为内底板纵骨腹板厚度; d_8 为外板纵骨腹板厚度。

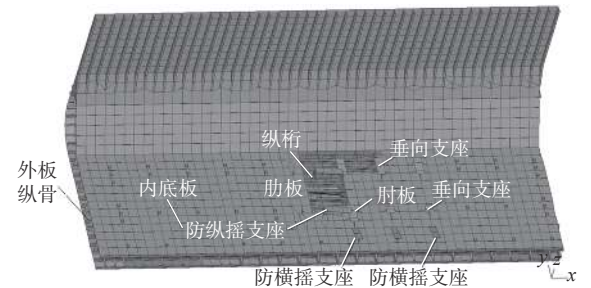


图 6 有限元计算模型
Fig. 6 Finite element model

根据船级社规范^[15]以及该船的设计方案, 选取本文研究对象的结构许用合成应力为 252 MPa。该结构的功能函数为

$$Z(X) = 252 - \sigma(X)_{\text{von}} \tag{11}$$

式中: $X = (X_1, X_2, \dots, X_{21})$, 为设计变量向量; $\sigma(X)_{\text{von}}$ 为通过有限元计算得到的实际合成应力。

MC 法和本文改进 RF-MC 法的计算结果对比如表 5 所示, 可以发现改进 RF-MC 法在只使用 250 个样本点计算结构失效概率的情况下, 计算值与理论值间的误差只有 2.77%, 精度较高。对于 A 型独立液舱支座这类复杂结构, 本文进行一次有限元计算需耗时约 2 min, 而 MC 法计算可靠度的样本数达到 10^5 量级, 由于本文方法使用较

表 4 A 型液舱支座结构的设计变量分布

Table 4 Design variables distribution of A-type liquid tank support

类别	编号	变量	参数1	参数2	分布类型
支座结构支反力	X ₁	f ₁ /N	5.24×10 ⁶	5.31×10 ⁶	均匀分布
	X ₂	f ₁ /N	4.72×10 ⁶	4.86×10 ⁶	均匀分布
	X ₃	f ₂ /N	2.57×10 ⁶	2.59×10 ⁶	均匀分布
	X ₄	f ₂ /N	2.74×10 ⁶	2.76×10 ⁶	均匀分布
	X ₅	f ₂ /N	1.94×10 ⁶	2.02×10 ⁶	均匀分布
	X ₆	f ₃ /N	2.09×10 ⁶	2.13×10 ⁶	均匀分布
	X ₇	f ₃ /N	1.98×10 ⁶	2.07×10 ⁶	均匀分布
	X ₈	f ₃ /N	2.36×10 ⁶	2.43×10 ⁶	均匀分布
	X ₉	f ₃ /N	2.81×10 ⁶	2.83×10 ⁶	均匀分布
	X ₁₀	f ₃ /N	3.97×10 ⁶	3.99×10 ⁶	均匀分布
	X ₁₁	f ₃ /N	4.87×10 ⁶	4.91×10 ⁶	均匀分布
板材厚度	X ₁₂	d ₁ /mm	14	1.4	正态分布
	X ₁₃	d ₁ /mm	12	1.2	正态分布
	X ₁₄	d ₂ /mm	18	1.8	正态分布
	X ₁₅	d ₃ /mm	18	1.8	正态分布
	X ₁₆	d ₄ /mm	11	1.1	正态分布
	X ₁₇	d ₄ /mm	11	1.1	正态分布
	X ₁₈	d ₅ /mm	18	1.8	正态分布
	X ₁₉	d ₆ /mm	14	1.4	正态分布
纵骨尺寸	X ₂₀	d ₇ /mm	12	1.2	正态分布
	X ₂₁	d ₈ /mm	12	1.2	正态分布

表 5 算例 3 计算结果

Table 5 Calculation results of Case 3

方法	失效概率P _f	相对误差/%	样本数量
MC	2.53×10 ⁻²	—	10 ⁵
改进RF-MC	2.46×10 ⁻²	2.77	250

少的样本点,可以使时间成本大为缩减。

4 结 论

综合考虑实际工程应用中对复杂结构进行可靠性分析的需求,本文提出了一种改进 RF-MC 法来计算结构失效概率。该方法通过 LOF 算法确定样本点的 $LOF_k(p)$ 筛选失效面附近的训练集,然后使用训练集构建 RF 模型拟合结构功能函数。不同于 BP-MC 法使用近似模型预测样本点到失效面间的距离,本文方法是通过选取合适的局部离群因子的上限阈值将样本集中远离失效面的噪声点予以排除,以保证模型拟合精度,在方法流程上更简便高效;另外,不同于 GBDT-MC 法使用 SMOTE 算法对失效面附近的样本进行插值,本文方法使用的样本都由 MC 法直接生成,从而

避免了插值样本可能不存在样本空间而造成模型拟合精度差的问题。

本文使用具有解析表达式的算例 1 并借助有限元软件求解的算例 2,验证了本文提出的改进 RF-MC 法在求解结构可靠性问题中的适用性。最后,将该方法应用于 LPG 船 A 型独立液舱支座结构的可靠性分析,结果表明,本文方法计算的结果与 MC 法的结果(工程中被普遍认为是理论解)之间的误差较小,且大幅减少了计算时间。本文的研究可为以后计算 LNG 船舶相似结构的失效概率提供高效、可靠的方法,也可以为机器学习方法在复杂船舶结构设计中的应用提供可行的技术方案。

参考文献:

[1] CHOI S K, CANFIELD R A, GRANDHI R V. Reliability-based structural design[M]. London: Springer, 2007.

[2] 孟广伟,李广博,李锋,等. 多项式基函数神经网络的结构可靠性分析 [J]. 北京航空航天大学学报, 2013, 39(11): 1460–1463.

MENG G W, LI G B, LI F, et al. Structural reliability analysis based on polynomial basis function neural network[J]. Journal of Beijing University of Aeronautics and Astronautics, 2013, 39(11): 1460–1463 (in Chinese).

[3] 金伟良,袁雪霞. 基于 LS-SVM 的结构可靠度响应面分析方法 [J]. 浙江大学学报(工学版), 2007, 41(1): 44–47,108.

JIN W L, YUAN X X. Response surface method based on LS-SVM for structural reliability analysis[J]. Journal of Zhejiang University (Engineering Science), 2007, 41(1): 44–47,108 (in Chinese).

[4] 李放,王德禹. 基于改进梯度提升决策树—蒙特卡罗法的超大型集装箱船绑扎桥可靠性分析 [J]. 中国舰船研究, 2020, 15(2): 63–69.

LI F, WANG D Y. Reliability analysis of lashing bridge of ultra-large container ship based on improved gradient boosting decision tree-Monte Carlo method[J]. Chinese Journal of Ship Research, 2020, 15(2): 63–69 (in Chinese).

[5] BREUNIG M M, KRIEGEL H P, NG R T, et al. LOF: identifying density-based local outliers[J]. ACM Sigmod Record, 2000, 29(2): 93–104.

[6] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5–32.

[7] BREIMAN L. Bagging predictors[J]. Machine Learning, 1996, 24(2): 123–140.

[8] 魏东海,王磊,赵志超,等. 随机森林算法在柴油机故障诊断中的应用 [J]. 机械设计与制造, 2020(7): 63–66.

WEI D H, WANG L, ZHAO Z C, et al. Application of random forest algorithm in fault diagnosis of diesel engine[J]. Machinery Design & Manufacture, 2020(7): 63–66 (in Chinese).